How well is the HoughLineFinder working?

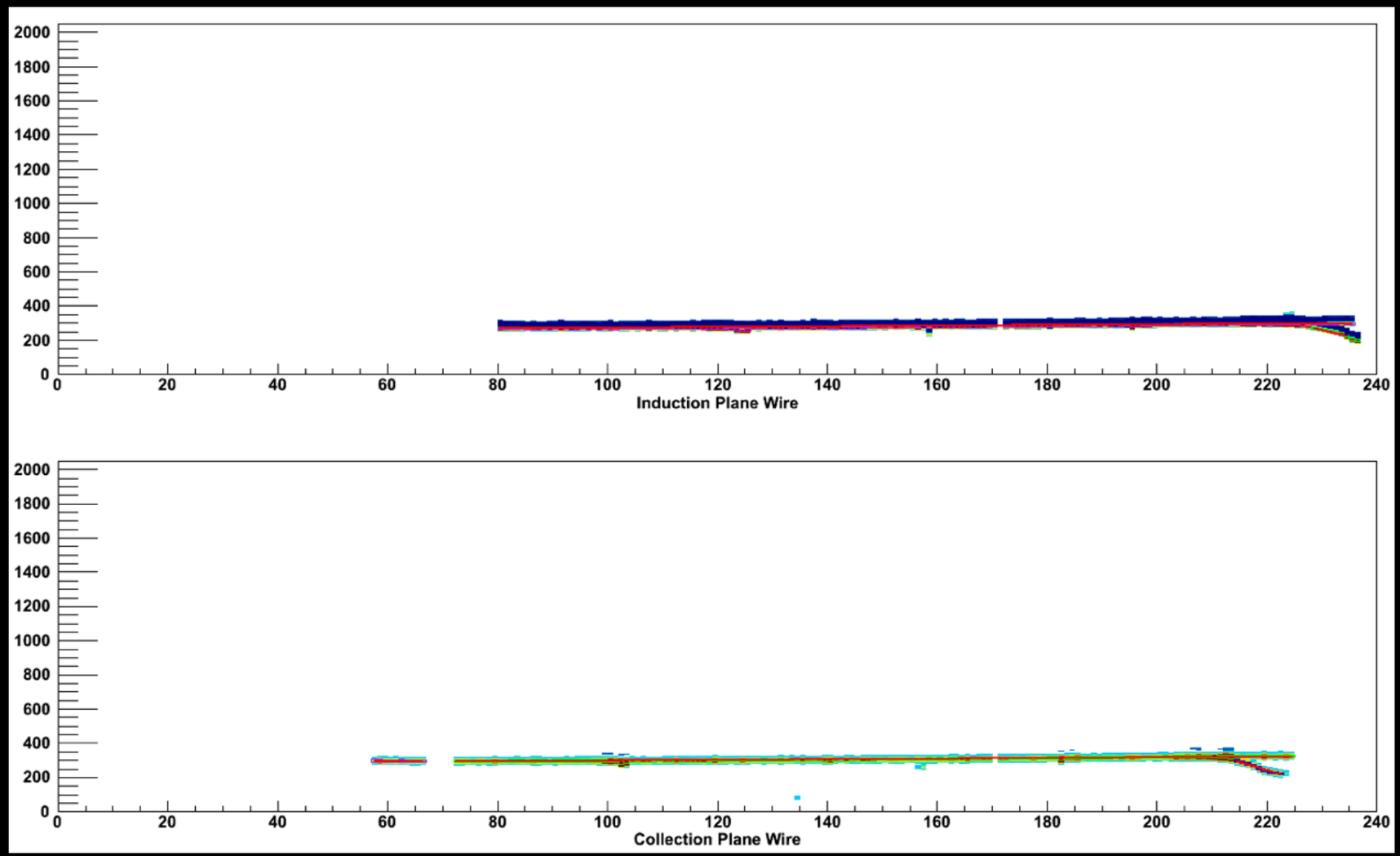
Joshua Spitz
8/19/2010

Algorithm tuning and improvement

- It is often difficult to quantitatively determine how well an individual algorithm in a reconstruction chain is working.
- The actual precision/proficiency/efficiency associated with the entire reconstruction chain is the number that eventually make it into the paper(s).
- However, developing a few quantitative "Figures of Merit" for tuning an individual algorithm's parameters and structure is important in ensuring that the final numbers are as strong as possible.
- We are now moving beyond "eye-balling" event displays in determining how well our algorithms are working.
  A good sign!

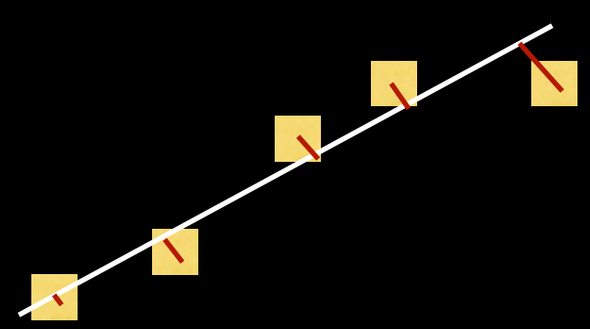
Reminder

- The HoughLineFinder algorithm takes hits that have been associated with DBSCAN clusters and finds line-like objects among the hit collections. Multiple line-like objects are allowed for each cluster. The algorithm returns the slope, intercept, and endpoints of each line.



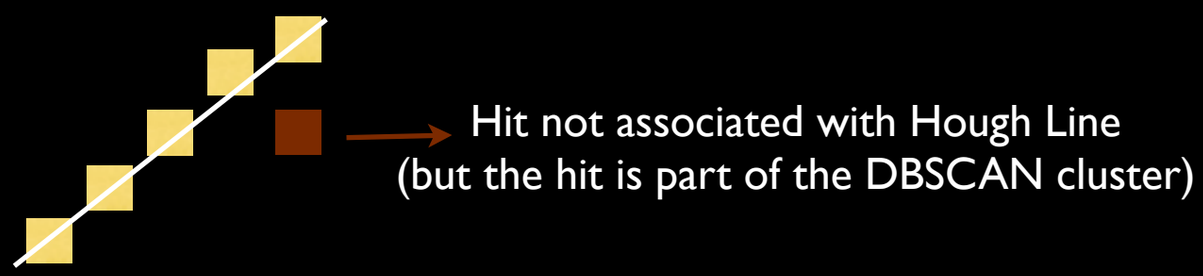
HoughLineFinder figures of merit

- Distance between a hit and a Hough line.
  - This is a "goodness-of-fit" measure. It can be turned into a chi^2, given a hit locational uncertainty. Note that I use the "distance of closest approach" definition.



HoughLineFinder figures of merit

- Number of DBSCAN cluster hits minus the number of Hough line hits.
- Fraction of total DBSCAN cluster hits associated with Hough lines.
  - These are measures of efficiency.
- Noise hits are not included in this study as they are only very rarely associated with DBSCAN clusters.



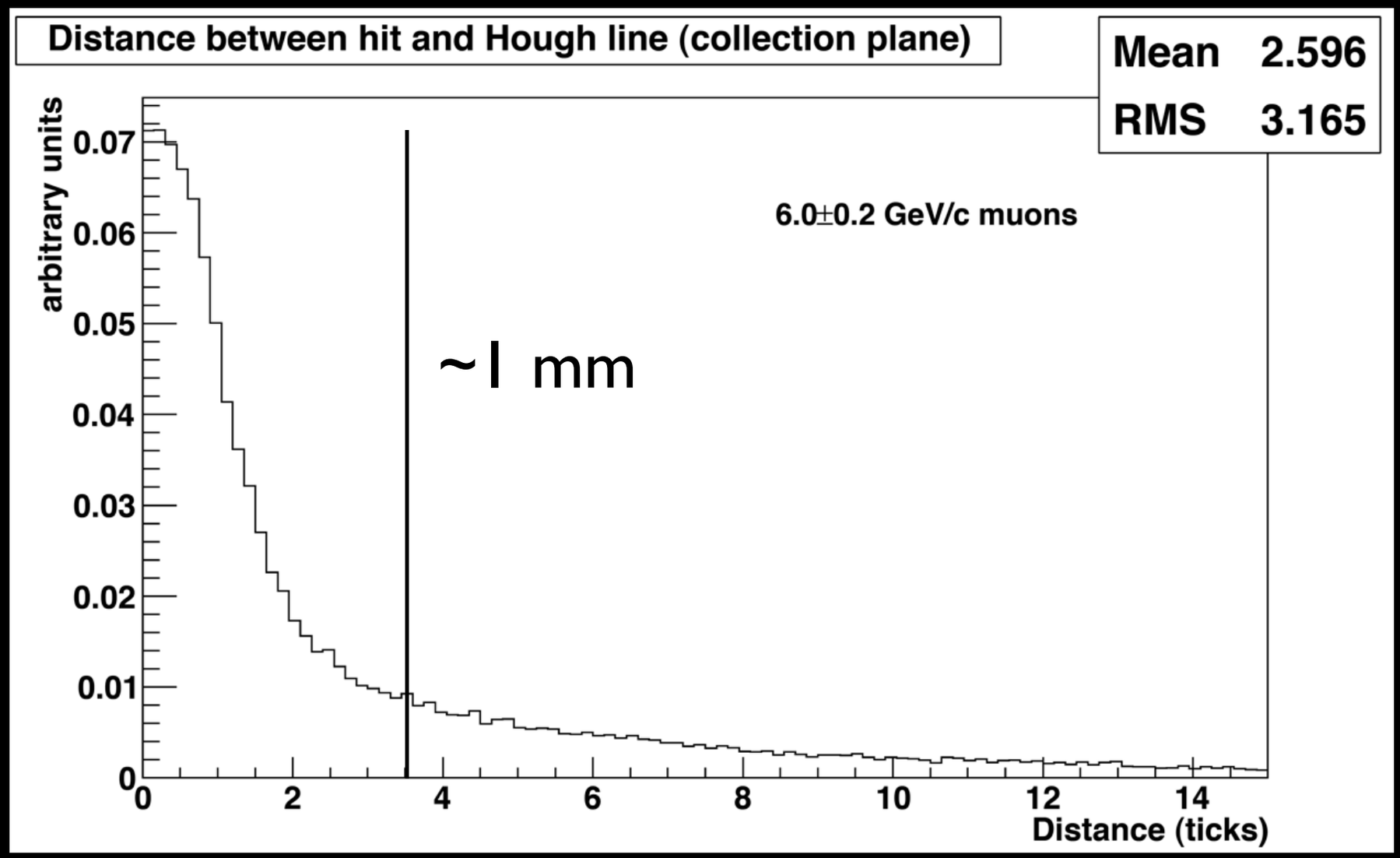
Simulation

- I have simulated 1000 6+/-.2 GeV/c mostly "flat" muons in ArgoNeuT in order to quantitatively determine how well the HoughLineFinder algorithm is working.

<param name="PDG"> <int> 13 </int> PDG codes of particles to simulate </param>
<param name="P0"> <float> 6.0 </float> Central momenta (GeV/c) of particles </par
<param name="SigmaP"> <float> 0.2 </float> Size of random variation in P0 </param>
<param name="PDist"> <int> 1 </int> How to vary the moment a (0 = uniform, 1 = gaussian) </param
<param name="X0"> <float> 0.0 </float> Central X position (cm) </param>
<param name="Y0"> <float> 0.0 </float> Central Y position (cm) </param>
<param name="Z0"> <float> 20.0 </float> Central Z position (cm) </param>
<param name="SigmaX"> <float> 30.0 </float> Random variation in x position </param>
<param name="SigmaY"> <float> 30.0 </float> Random variation in y position </param>
<param name="SigmaZ"> <float> 0.0 </float> Random variation in z position </param>
<param name="PosDist"> <int> 0 </int> How to vary xyz postions (0 = uniform, 1 = gaussian) </param>
<param name="Theta0XZ"> <float> 0 </float> Angles in XZ plane (degrees) </param>
<param name="Theta0YZ"> <float> -3.3 </float> Angles in YZ plane (degrees) </param>
<param name="SigmaThetaXZ"> <float> 0.2 </float> Variation in XZ angle (degrees) </param>
<param name="SigmaThetaYZ"> <float> 0.2 </float> Variation in YZ angle (degrees) </param>
<param name="AngleDist"> <int> 1 </int> How to vary angles (0=uniform, 1=gaussian) </param>

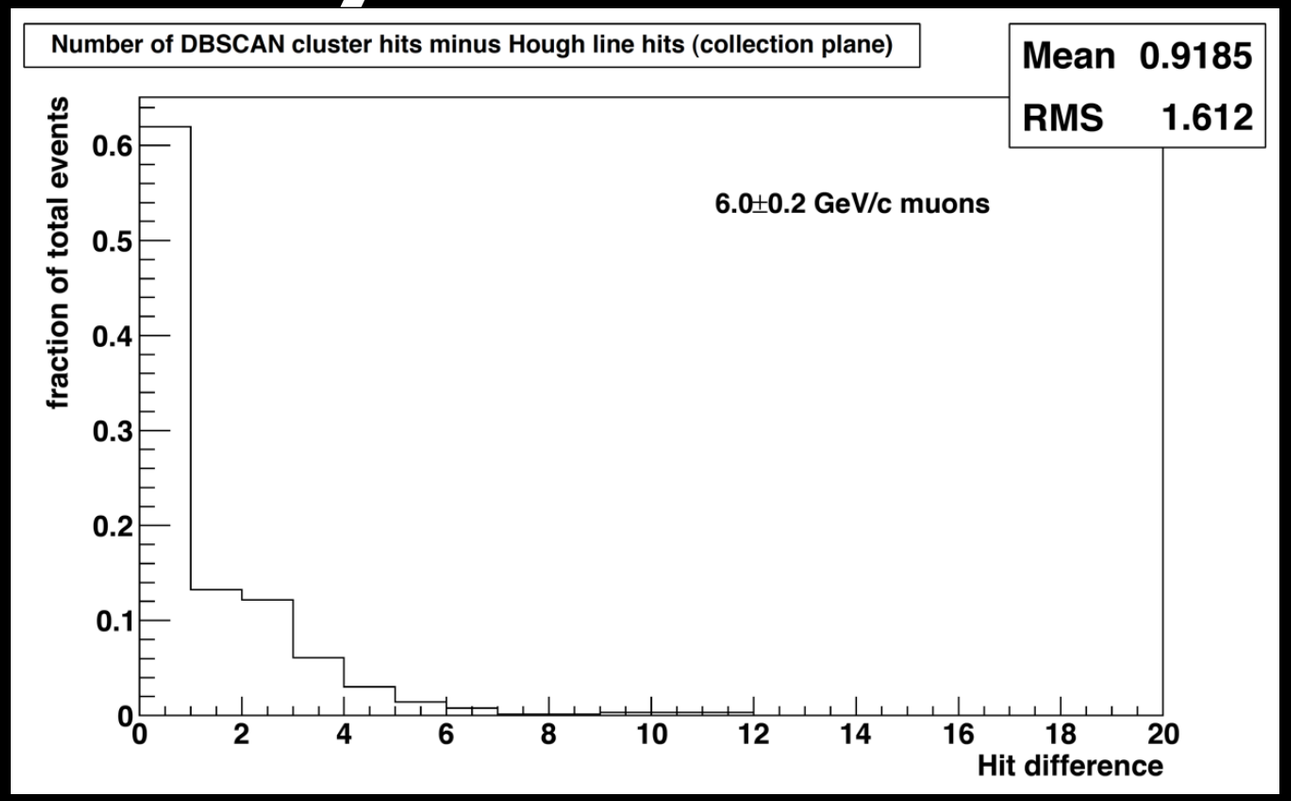
Goodness-of-fit

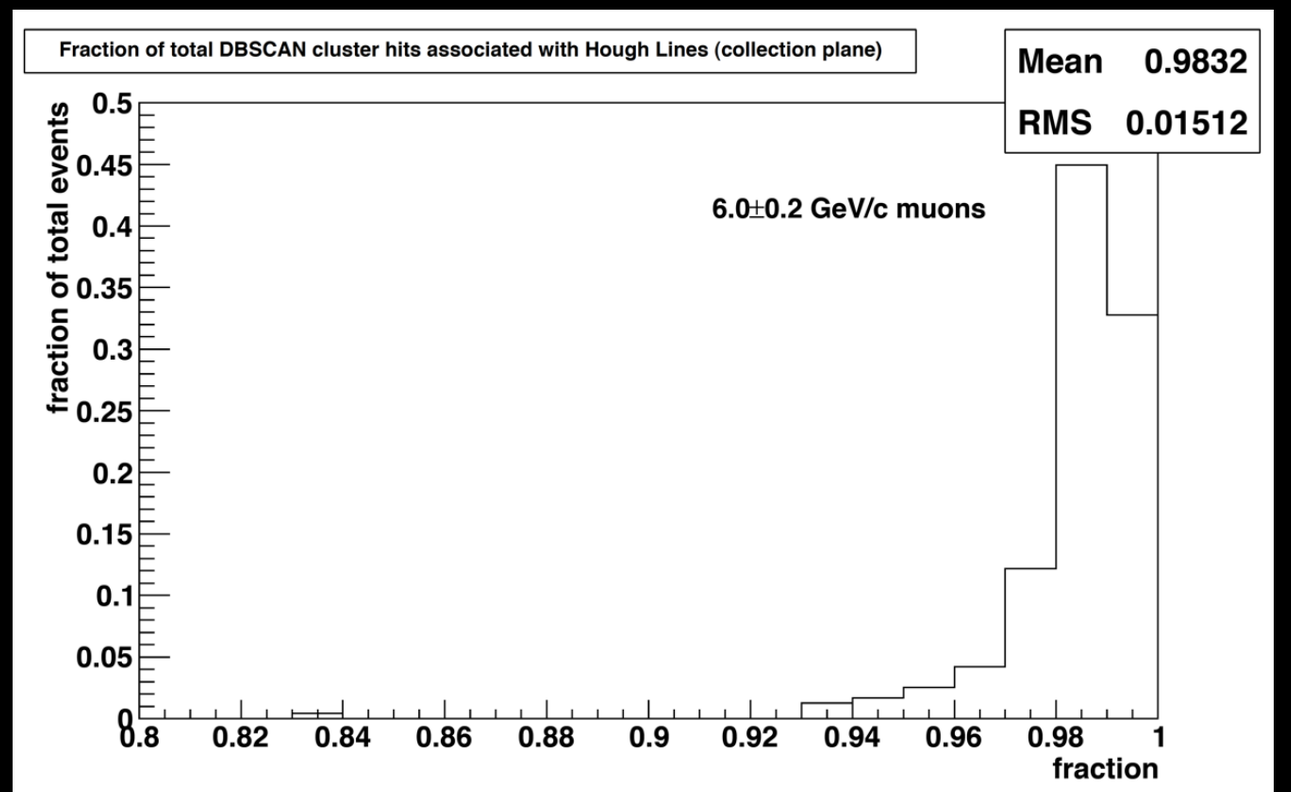
- The distance of closest approach between a hit and its Hough line for all hits in 1000 simulated muons



Efficiency

- Recall that the HoughLineFinder's input is DBSCAN cluster hits.
- These plots compare the number of hits associated with all DBSCAN clusters to the number of hits associated with all Hough lines in muon (line-like) events.
- As muons are largely "line-like", we expect the number of hits for DBSCAN and HoughLineFinder to be approximately equal.





Tuning

- The figures of merit are used to tune the following HoughLineFinder parameters:
  - Max number of lines that can be found per cluster.
  - Maximum Distance of hit from line, in consideration of hit width.
  - Number of rho and theta cells in the Hough Accumulator. These numbers are proportional to the resolution of the parameters of a fit line. Also, they set the requirements for breaking a line.

"Breaking a line". Is this set of hits two lines or one?

Conclusion

- A few quantitative checks of the HoughLineFinder's abilities have been developed.
- These checks can be (and have been) used to fine-tune the algorithm's parameters.
- There are a few more plots to make:
  - Simulating muons at various angles.
  - 2d histo: # of hits in a line vs. average distance between hits and line.